

Natural Language Syntax

Formele en Natuurlijke Talen
Lecture 8

Goals for today

- **Understand** the fundamental properties of natural language syntax
- **Establish** how these properties do(n't) vary across languages
- **Model** these properties using formal grammars (Test case: a fragment of Dutch)

Based on slides by Rick Nouwen and Jakub Dotlačil

What does it mean to 'know' a language?

Principles: competence

What does it mean to 'know' a language?

Colourless green ideas sleep furiously

*Ideas green sleep colourless furiously

Principles: competence

What does it mean to 'know' a language?

Colourless green ideas sleep furiously

Well-formed but nonsensical

*Ideas green sleep colourless furiously

Principles: competence

What does it mean to 'know' a language?

Colourless green ideas sleep furiously

Well-formed but nonsensical

*Ideas green sleep colourless furiously

Ill-formed but nonsensical

Principles: competence

What does it mean to 'know' a language?

Colourless green ideas sleep furiously

Well-formed but nonsensical

*Ideas green sleep colourless furiously

Ill-formed but nonsensical

Noam Chomsky, *Syntactic Structures* (1957)

Principles: ambiguity

Language is rife with **ambiguity** (one form, multiple possible interpretations)

Principles: ambiguity

Language is rife with **ambiguity** (one form, multiple possible interpretations)

Lexical ambiguity: Ik zie een **bank**.

Principles: ambiguity

Language is rife with **ambiguity** (one form, multiple possible interpretations)

Lexical ambiguity: Ik zie een **bank**.



Principles: ambiguity

Language is rife with **ambiguity** (one form, multiple possible interpretations)

Lexical ambiguity: Ik zie een **bank**.



Another source of ambiguity

Some phrases are ambiguous without lexical ambiguity:

Another source of ambiguity

Some phrases are ambiguous without lexical ambiguity:



Source: @kenzianidiot on Twitter

Structural ambiguity

Where does this ambiguity come from?

Structural ambiguity

Where does this ambiguity come from?

Dit is de vrouw die Jan kuste.

Sue zag de man met de verrekijker.

Nederlandse kaasliefhebber

These phrases have more than one possible interpretation

Structural ambiguity

Where does this ambiguity come from?

Dit is de vrouw die Jan kuste.

Sue zag de man met de verrekijker.

Nederlandse kaasliefhebber

These phrases have more than one possible interpretation

No lexical ambiguity \Rightarrow the ambiguity must come from elsewhere

Structural ambiguity

Where does this ambiguity come from?

Dit is de vrouw die Jan kuste.

Sue zag de man met de verrekijker.

Nederlandse kaasliefhebber

These phrases have more than one possible interpretation

No lexical ambiguity \Rightarrow the ambiguity must come from elsewhere

Nederlandse kaasliefhebber

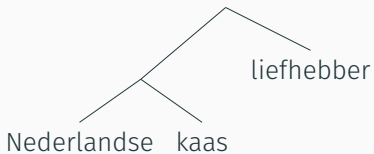
Structural ambiguity

Nederlandse kaasliefhebber



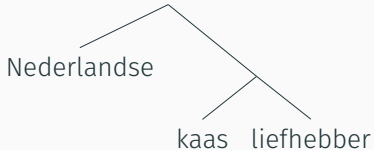
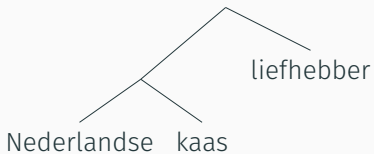
Structural ambiguity

Nederlandse kaasliefhebber



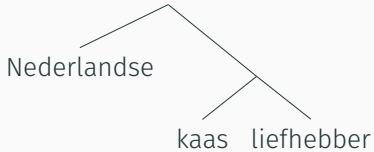
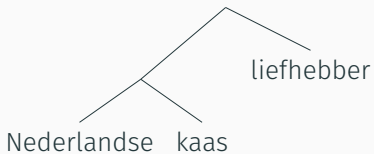
Structural ambiguity

Nederlandse kaasliefhebber



Structural ambiguity

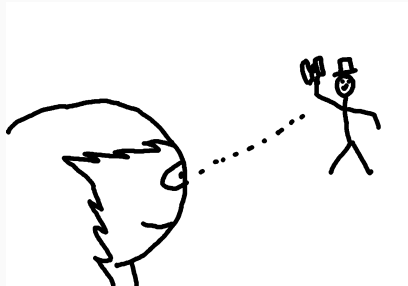
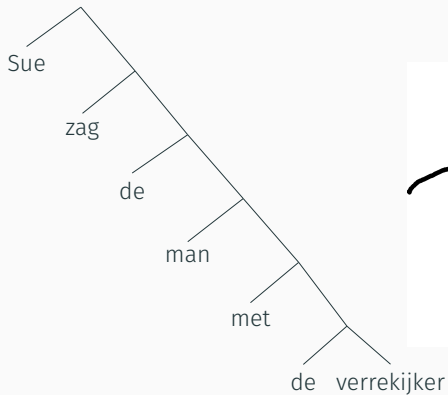
Nederlandse kaasliefhebber



Sue zag de man met de verrekijker.

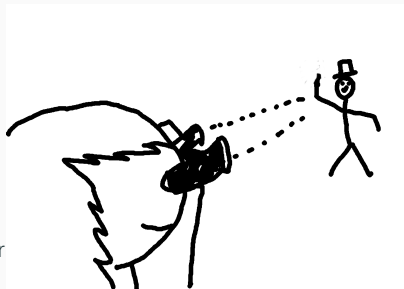
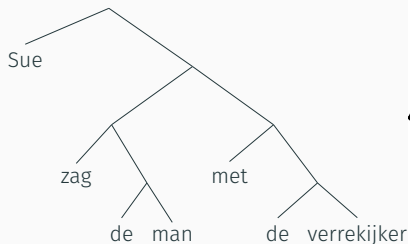
Structural ambiguity

Sue zag de man met de verrekijker.



Structural ambiguity

Sue zag de man met de verrekijker.



Constituency

Constituent: A group of words that together form a unit

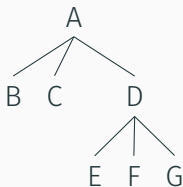
In trees: constituent = one node and every node it **dominates**

Constituency

Constituent: A group of words that together form a unit

In trees: constituent = one node and every node it **dominates**

α dominates β iff $\alpha \neq \beta$ and we can trace a path from α to β without going up

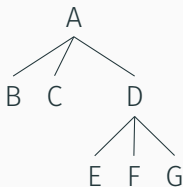


Constituency

Constituent: A group of words that together form a unit

In trees: constituent = one node and every node it **dominates**

α dominates β iff $\alpha \neq \beta$ and we can trace a path from α to β without going up



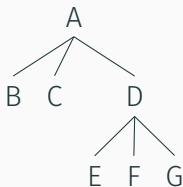
D dominates E, F, G

Constituency

Constituent: A group of words that together form a unit

In trees: constituent = one node and every node it **dominates**

α dominates β iff $\alpha \neq \beta$ and we can trace a path from α to β without going up



D dominates E, F, G

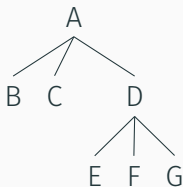
A dominates B, C, D, E, F, G

Constituency

Constituent: A group of words that together form a unit

In trees: constituent = one node and every node it **dominates**

α dominates β iff $\alpha \neq \beta$ and we can trace a path from α to β without going up



D dominates E, F, G

A dominates B, C, D, E, F, G

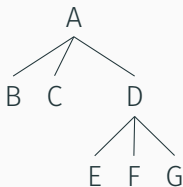
C dominates nothing

Constituency

Constituent: A group of words that together form a unit

In trees: constituent = one node and every node it **dominates**

α dominates β iff $\alpha \neq \beta$ and we can trace a path from α to β without going up



D dominates E, F, G

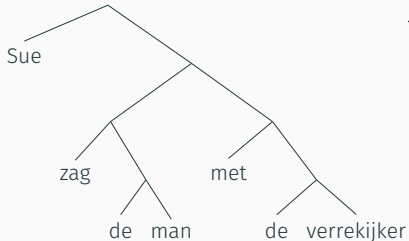
A dominates B, C, D, E, F, G

C dominates nothing

B and E **do not** form a constituent

D, E, F, G **do** form a constituent

Question 1



Which of the following strings corresponds to a constituent in this tree?

- (a) Sue zag de man
- (b) verrekijker
- (c) Sue zag de man met de verrekijker
- (d) de man met de verrekijker

Question 1

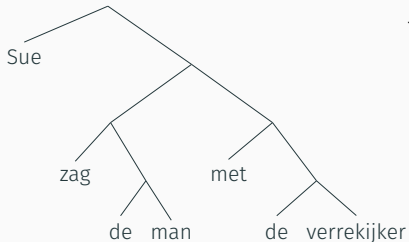


**1** Ga naar woelap.com

**2** Voer de code van het evenement in de bovenste banner in

Evenementcode
FENT08

Antwoorden per 100 inschelden



Which of the following strings corresponds to a constituent in this tree?

- (a) Sue zag de man
- (b) verrekijker**
- (c) Sue zag de man met de verrekijker**
- (d) de man met de verrekijker

Constituency in Language

[Sue [kuste Jan]]

but not [[Sue kuste] Jan]

Constituency in Language

[Sue [kuste Jan]]

but not [[Sue kuste] Jan]

Tests for constituency (not 100% reliable, but a good start):

Substitution test: Only constituents can be replaced by pronouns

- (1) a. Sue **kuste Jan**. Ze **deed dat** vorige week.
 b. Sue kuste Jan. *Hij was dat vorige week.

Constituency in Language

[Sue [kuste Jan]]

but not [[Sue kuste] Jan]

Tests for constituency (not 100% reliable, but a good start):

Substitution test: Only constituents can be replaced by pronouns

- (1) a. Sue **kuste Jan**. Ze **deed dat** vorige week.
b. Sue kuste Jan. *Hij was dat vorige week.

- (2) **The spy with the beret** waltzed. → **She** waltzed.

Diagnosing constituents

Diagnosing constituents

Question test: only constituents make good answers

- (3) a. Who came to the party? The lady with the beret.
What did the lady come to the party with? The beret.

Diagnosing constituents

Question test: only constituents make good answers

- (3) a. Who came to the party? The lady with the beret.
What did the lady come to the party with? The beret.

Constituency is built recursively:

- (4) a. Where did you send the money to?
[A bank in Switzerland].
- b. Where did you send the money?
[To [a bank in Switzerland]].
- c. What did you do?
[Send the money [to [a bank in Switzerland]]].

What's necessary in natural language
syntax

Basic anatomy of a sentence

Basic ingredients of a sentence:

- Subject (*onderwerp*): S
- Verb: V
- (Possibly) an object (*lijdend voorwerp*): O

Basic anatomy of a sentence

Basic ingredients of a sentence:

- Subject (*onderwerp*): S
- Verb: V
- (Possibly) an object (*lijdend voorwerp*): O

John saw Sue.

Jan kuste Piet.

S V O

Basic anatomy of a sentence

Basic ingredients of a sentence:

- Subject (*onderwerp*): **S**
- Verb: **V**
- (Possibly) an object (*lijdend voorwerp*): **O**


The diagram illustrates the basic anatomy of the sentence "de man die Piet geslagen heeft haat mensen met aggresieve honden." The words are color-coded: "de man die Piet geslagen heeft" is in red, "haat" is in black, and "mensen met aggresieve honden." is in red. A bracket labeled "S" (Subject) is placed under the red text "de man die Piet geslagen heeft". A bracket labeled "V" (Verb) is placed over the black word "haat". A bracket labeled "O" (Object) is placed under the red text "mensen met aggresieve honden.".

Basic word order

(5) sho-nen-ga boru-o kero
boy ball kick
'The boy kicks the ball.'

(Japanese)

Basic word order

- (5) sho-nen-ga boru-o kero
boy ball kick
'The boy kicks the ball.'

(Japanese)

- (6) De jongen trapt de bal. (SVO)
- (7) De jongen de bal trapt. (SOV)
- (8) Trapt de jongen de bal. (VSO)
- (9) Trapt de bal de jongen. (VOS)
- (10) De bal trapt de jongen. (OVS)
- (11) De bal de jongen trapt. (OSV)

Basic word order

(5) sho-nen-ga boru-o kero (Japanese)
 boy ball kick
 'The boy kicks the ball.'

- SVO \pm 41% (e.g. English, Mandarin)
- SOV: \pm 47% (e.g. Japanese, Turkish)
- VSO: \pm 8% (e.g. Tagalog, Irish)
- VOS: \pm 2.5% (e.g. Fijian, Ojibwe (Algonquian; US/Canada))
- OVS: \pm 0.8% (e.g. Urarina (isolate; Peru))
- OSV: \pm 0.4% (e.g. Warao (isolate; Venezuela/Guyana/Suriname))

Dryer (2013). Order of Subject, Object and Verb. In Dryer & Haspelmath (Eds.): *The World Atlas of Language Structures Online*.

Basic word order: Dutch

Jan kuste Piet.

So is Dutch SVO?

Basic word order: Dutch

Jan kuste Piet.

So is Dutch SVO?

Some complications:

Ik geloof dat [Jan Piet kuste]. (SV-SOV)

Basic word order: Dutch

Jan kuste Piet.

So is Dutch SVO?

Some complications:

Ik geloof dat [Jan Piet kuste]. (SV-SOV)

Gisteren at Jan sushi. (VSO)

Dutch's basic word order is SOV, but exhibits so-called **V2** ('verb second'):

- ★ In unembedded sentences (*hoofdzinnen*) there is always exactly one constituent before the (inflected) verb.

Dutch's basic word order is SOV, but exhibits so-called **V2** ('verb second'):

- ★ In unembedded sentences (*hoofdzinnen*) there is always exactly one constituent before the (inflected) verb.

[Jan] **at** sushi. (V2)

V2 in Dutch

Dutch's basic word order is SOV, but exhibits so-called **V2** ('verb second'):

- ★ In unembedded sentences (*hoofdzinnen*) there is always exactly one constituent before the (inflected) verb.

[Jan] **at** sushi. (V2)

[Gisteren] **at** Jan sushi. (V2)

[Op de dag dat Jan Piet zag], **zag** Piet Jan niet. (V2)

V2 in Dutch

But: why say that Dutch is SOV and there is a V2 rule?

V2 in Dutch

But: why say that Dutch is SOV and there is a V2 rule?

Why not the other way around: Dutch is SVO and there is a V-last rule in embedded clauses (*bijzinnen*)?

Jan at sushi./Ik geloof dat Jan sushi at.

V2 in Dutch

But: why say that Dutch is SOV and there is a V2 rule?

Why not the other way around: Dutch is SVO and there is a V-last rule in embedded clauses (*bijzinnen*)?

Jan at sushi./Ik geloof dat Jan sushi at.

Reason to reject this view: SOV occurs across many more contexts \Rightarrow **basic**:

Jan sushi eten?

Jan gaat sushi etend door het leven.

Jan heeft sushi gegeten.

V2 in Dutch

But: why say that Dutch is SOV and there is a V2 rule?

Why not the other way around: Dutch is SVO and there is a V-last rule in embedded clauses (*bijzinnen*)?

Jan at sushi./Ik geloof dat Jan sushi at.

Reason to reject this view: SOV occurs across many more contexts ⇒ **basic**:

Jan sushi eten?

Jan gaat sushi etend door het leven.

Jan heeft sushi gegeten.

3-4 year old children:

koekje eten

eat cookie

(Dutch; OV)

(English; VO) 17

V2 in Dutch as a diagnostic

The V2 property of Dutch can be used to **identify** constituents:

V2 in Dutch as a diagnostic

The V2 property of Dutch can be used to **identify** constituents:

V2 in Dutch as a diagnostic

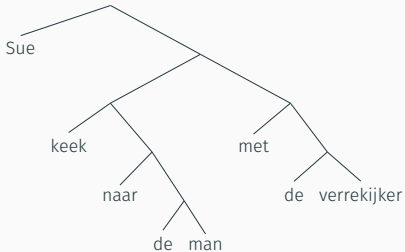
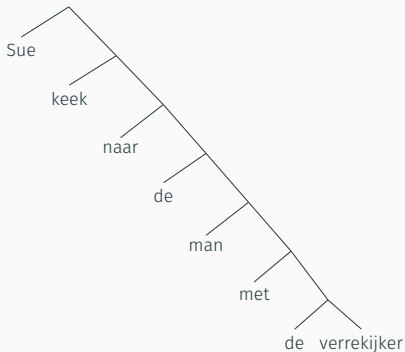
The V2 property of Dutch can be used to **identify** constituents:

Sue keek naar de man met de verrekijker.

V2 in Dutch as a diagnostic

The V2 property of Dutch can be used to **identify** constituents:

Sue keek naar de man met de verrekijker.



V2 in Dutch as a diagnostic

The V2 property of Dutch can be used to **identify** constituents:

Sue keek naar de man met de verrekijker.

[Naar de man] keek Sue met de verrekijker.

[Naar de man met de verrekijker] keek Sue (niet Piet).

How constituents are put together

Lexical categories

Words fall into different categories ('lexical classes') which affect how they can be combined:

- **Nouns:** *walrus, tractor, nihilism, ...*
- **Verbs:** *run, electrify, think, ...*
- **Adjectives:** *dusty, tall, Australian, ...*
- **Adverbs:** *well, quickly, often, ...*
- **Prepositions:** *to, in, between, ...*
- ...

Lexical categories

Words fall into different categories ('lexical classes') which affect how they can be combined:

- **Nouns:** *walrus, tractor, nihilism, ...*
- **Verbs:** *run, electrify, think, ...*
- **Adjectives:** *dusty, tall, Australian, ...*
- **Adverbs:** *well, quickly, often, ...*
- **Prepositions:** *to, in, between, ...*
- ...

Membership in a category restricts (and arguably is defined by) with what other kinds of objects a word can be put together

Heads and phrases

The **head** (*hoofd*)
is the **nucleus** of a
constituent.

Heads and phrases

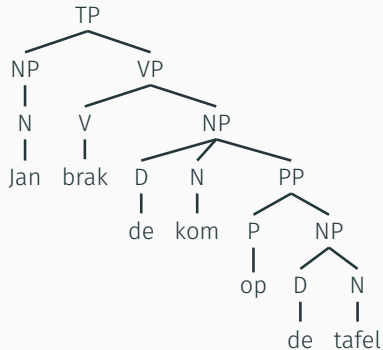
The **head** (*hoofd*)
is the **nucleus** of a
constituent.

Nouns are heads
of noun phrases
(NP), verbs of verb
phrases (VP), etc.

Heads and phrases

The **head** (*hoofd*) is the **nucleus** of a constituent.

Nouns are heads of noun phrases (NP), verbs of verb phrases (VP), etc.



Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

wijn
vin

Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

witte wijn
vin blanc

Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

droge witte wijn
 vin blanc sec

Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

heerlijke droge witte wijn
 vin blanc sec délicieux

Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

heerlijke droge witte wijn
 vin blanc sec délicieux

Head-initial: verb before object, noun before adjectives, prepositions...

Headedness

Typically heads go at the edge of a phrase, but which edge can vary by (or even within a single) language!

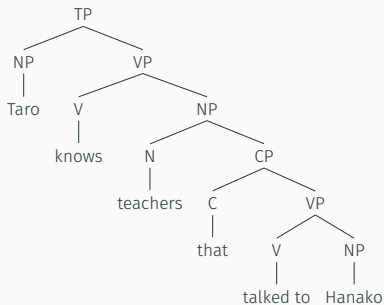
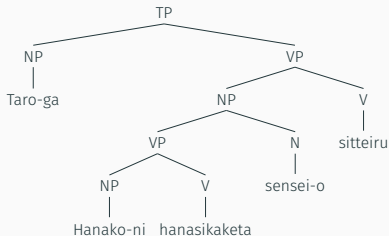
heerlijke droge witte wijn
 vin blanc sec délicieux

Head-initial: verb before object, noun before adjectives, prepositions...

Head-final: verb after object, noun after adjectives, postpositions...

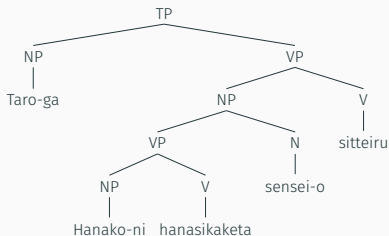
Head-initial vs. head-final

- (6) Taro-ga Hanako-ni hanasikaketa sensei-o sitteiru
Taro Hanako talked-to teacher knows
'Taro knows teachers that talked to Hanako.'

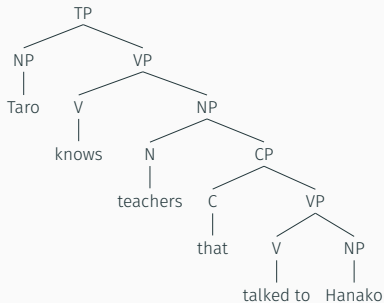


Head-initial vs. head-final

- (6) Taro-ga Hanako-ni hanasikaketa sensei-o sitteiru
Taro Hanako talked-to teacher knows
'Taro knows teachers that talked to Hanako.'



Japanese – head-final



English – head-initial

Applying CFGs to natural language

How can we model the constituency structure of languages?

Formal grammars and natural language

How can we model the constituency structure of languages?

For now: *formal grammars* (in particular **context-free grammars**)

Intuitive correspondences to parts of a CFG $\langle \Sigma, N, S, P \rangle$

- Σ : words of the language

Formal grammars and natural language

How can we model the constituency structure of languages?

For now: *formal grammars* (in particular **context-free grammars**)

Intuitive correspondences to parts of a CFG $\langle \Sigma, N, S, P \rangle$

- Σ : words of the language
- N : Labels for categories (N, V, ...) and phrases (NP, VP, ...)

Formal grammars and natural language

How can we model the constituency structure of languages?

For now: *formal grammars* (in particular **context-free grammars**)

Intuitive correspondences to parts of a CFG $\langle \Sigma, N, S, P \rangle$

- Σ : words of the language
- N : Labels for categories (N, V, ...) and phrases (NP, VP, ...)
- S : Label for sentences

Formal grammars and natural language

How can we model the constituency structure of languages?

For now: *formal grammars* (in particular **context-free grammars**)

Intuitive correspondences to parts of a CFG $\langle \Sigma, N, S, P \rangle$

- Σ : words of the language
- N : Labels for categories (N, V, ...) and phrases (NP, VP, ...)
- S : Label for sentences
- P : 'Phrase structure rules'
 - $X \rightarrow \dots$: tells you what words are in category X
 - $XP \rightarrow \dots$: recipe for how to build an XP

Formal grammars and natural language

How can we model the constituency structure of languages?

For now: *formal grammars* (in particular **context-free grammars**)

Intuitive correspondences to parts of a CFG $\langle \Sigma, N, S, P \rangle$

- Σ : words of the language
 - N : Labels for categories (N, V, ...) and phrases (NP, VP, ...)
 - S : Label for sentences
 - P : 'Phrase structure rules'
 - $X \rightarrow \dots$: tells you what words are in category X
 - $XP \rightarrow \dots$: recipe for how to build an XP
- \Rightarrow Each application of a rule forms a constituent

A formal grammar for Dutch NPs

A formal grammar for Dutch NPs

$\Sigma = \{\text{de, man, bejaarde, gemene, enorm, onwijs}\}$

$\left\{ \begin{array}{l} \text{de bejaarde man} \\ \text{de man} \\ \text{de enorm gemene man} \\ \text{de onwijs gemene man} \\ \text{de onwijs bejaarde gemene man} \\ \text{de gemene bejaarde man} \\ \text{de bejaarde gemene man} \\ \dots \end{array} \right\}$

*de onwijs man

*de enorm man

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D N$

$\text{NP} \rightarrow D A N$

$\text{NP} \rightarrow D \text{Adv} A N$

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D (\text{Adv } A) N$

$\text{NP} \rightarrow D A N$

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D (\text{Adv } A) N$

$\text{NP} \rightarrow D A N$

Does this grammar overgenerate? (Produce strings that it shouldn't)

Undergenerate? (Fail to produce strings it should)

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D (\text{Adv } A) N$

$\text{NP} \rightarrow D A N$



1 Ga naar wooclap.com

2 Voer de code van het evenement in de bovenste banner in

Evenementcode
FENT08

Antwoorden per sms inschakelen

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

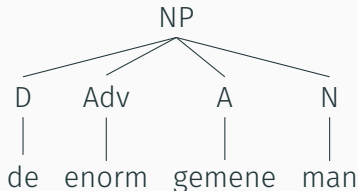
$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D (\text{Adv } A) N$

$\text{NP} \rightarrow D A N$



A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D (\text{Adv } A) N$

$\text{NP} \rightarrow D A N$

de gemene bejaarde man

undergeneration!

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A \text{ NP}$

$\text{NP} \rightarrow \text{Adv} \text{ NP}$

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D \text{ NP} \quad \star \quad \text{Recursive production rule}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A \text{ NP} \quad \star \quad \text{Recursive production rule}$

$\text{NP} \rightarrow \text{Adv} \text{ NP} \quad \star \quad \text{Recursive production rule}$

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

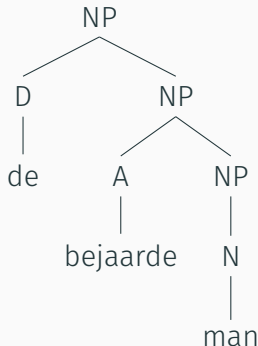
$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A \text{ NP}$

$\text{NP} \rightarrow \text{Adv NP}$



A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

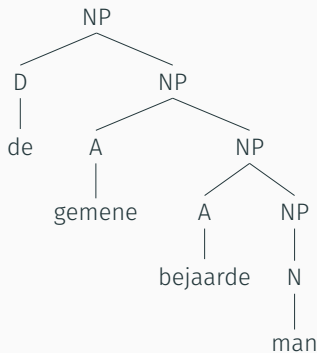
$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A \text{ NP}$

$\text{NP} \rightarrow \text{Adv NP}$



A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

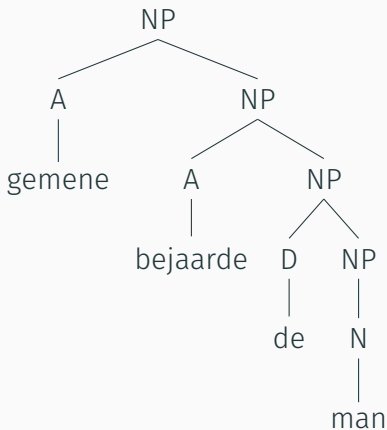
$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{NP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A \text{ NP}$

$\text{NP} \rightarrow \text{Adv NP}$



overgeneration!

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

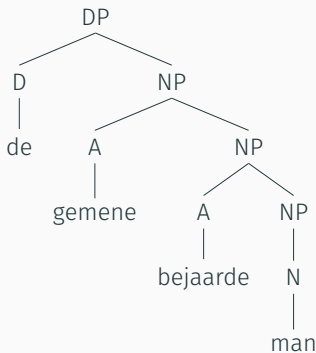
$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{DP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A \text{ NP}$

$\text{NP} \rightarrow \text{Adv} \text{ NP}$



A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

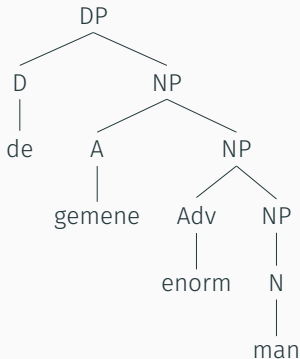
$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{DP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A \text{ NP}$

$\text{NP} \rightarrow \text{Adv} \text{ NP}$



over-generation!

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{DP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow A^P \text{ NP}$

$A^P \rightarrow (\text{Adv}) A$

A formal grammar for Dutch NPs

{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$D \rightarrow \text{de}$

$N \rightarrow \text{man}$

$A \rightarrow \text{bejaarde} \mid \text{gemene}$

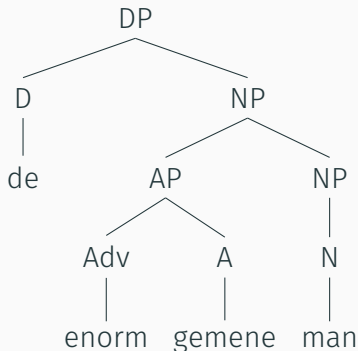
$\text{Adv} \rightarrow \text{onwijs} \mid \text{enorm}$

$\text{DP} \rightarrow D \text{ NP}$

$\text{NP} \rightarrow N$

$\text{NP} \rightarrow \text{AP} \text{ NP}$

$\text{AP} \rightarrow (\text{Adv}) A$



A formal grammar for Dutch NPs

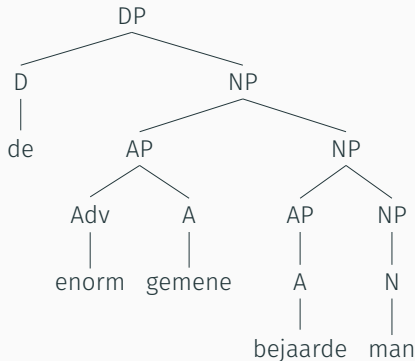
{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$DP \rightarrow D \ NP$

$NP \rightarrow N$

$NP \rightarrow AP \ NP$

$AP \rightarrow (Adv) \ A$



A formal grammar for Dutch NPs

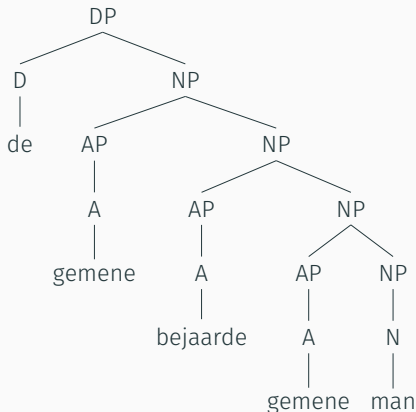
{de bejaarde man, de man, de enorm gemene man, de onwijs gemene man, de onwijs bejaarde gemene man, de gemene bejaarde man, de bejaarde gemene man, ...}

$DP \rightarrow D \ NP$

$NP \rightarrow N$

$NP \rightarrow AP \ NP$

$AP \rightarrow (Adv) \ A$



- Automatic parsers:
Alpino: <https://www.let.rug.nl/vannoord/alp/Alpino/>
Berkeley neural parser:
<https://github.com/nikitakit/self-attentive-parser>
- Corpora (treebanks): e.g. Penn Treebank
- Machine translation, information extraction

The queen spoke with the Dutch prime minister in UK.

Summary

- Concepts important in natural language syntax:
 - Headedness
 - Word order (SVO, VOS,...)
 - Constituency
 - Ambiguity
- We can analyze natural language syntax using formal grammars.
- Question to think about: can context-free grammars capture *all* linguistic structures?